

서울시 청년 1인 가구
임대주택
청약 예측 서비스

PPT PRESENTATION



목차

1. 기획 의도 / 개발 목표
2. 기술 Stacks
3. 개발 스케줄표
4. 빅데이터 분석 결과 보고서
5. 요구사항 정의서
6. 화면 설계서
7. UML

PPT PRESENTATION

20~30대 1인 가구, 시세보다 2~30% 저렴한 임대주택 인기

- 2030 청년 1인 가구 비율은 점점 늘어 가고 있으나
치솟는 집값으로 인한 주거비 부담과 전세사기 피해 확산으로
임대주택에 대한 관심이 늘고 있음

1인 가구 추이 (단위: 만 가구)

자료: 통계청



1인가구 소득분포

단위: %, 지난해 기준



자료: 통계청

The JoongAng

구분	보증금, 월세 등 주거비 보조	주택 관련 대출 및 이자 지원	공임대주택 공급	자기 소유 지원	주택 개량 및 보수 지원	정보 제공 및 상담
남성	65.8	67.1	67.0	67.2	62.3	63.1
20	70.9	71.1	71.0	71.8	64.7	65.6
30	66.9	69.0	67.7	68.6	63.6	64.8
40	64.3	66.9	68.8	66.6	62.0	63.8
50	64.2	64.6	65.0	66.3	61.0	62.1
여성	66.3	66.5	67.0	66.2	63.4	63.1
20	70.6	68.0	68.9	69.5	64.3	65.5
30	70.1	70.7	70.9	68.4	66.0	64.4
40	65.7	69.0	66.3	67.3	62.6	64.9
50	62.1	65.1	65.3	65.0	61.7	64.4

*출처: 서울시 1인가구 실태조사 및 제도개선 연구용역 보고서

D 대한경제 | 2023.04.13.

[마피 사기주의보] 전세사기에 임대주택으로 몰리는 세입자들

빌라왕 전세사기 여파로 인해 2030세대 세입자들이 보증금 리스크가 낮은 임대주택 등을 주목하고 있다... 합리적인 임대료로 주거 안정이 보장되는 민간임대주택이 내 집 마련 대안으로 급부상하면서 높은 인기를 끌고...

EKN 에너지경제 | 2022.10.06.

2030 영끌족은 옛말...LH 청년 매입임대주택 '인기'

전세사기·강통전세에 대한 불안 등도 영향을 준 것을 분석된다. 매입임대는 LH에서 주도하는 주택 제도로 안전하다는 인식이 있다. 특히 청년매입임대주택은 최장 6...

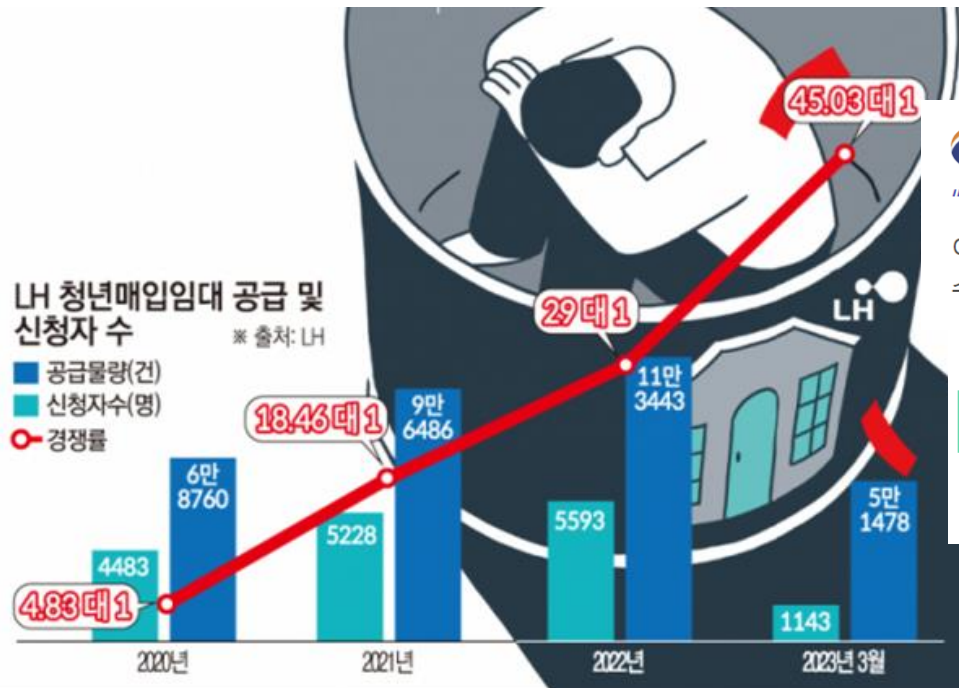
K 국제신문 PICK | 10면 1단 | 2023.07.27. | 네이버뉴스

전세사기 걱정없는 LH행복주택 인기

행복주택의 높아진 인기를 실감케 했다. 부산 모라는 100가구 모집에 246명 지원 (경쟁률 2.5 대 1)했고, 울산 송정은 220가구 모집에 531명(2.4 대 1)이 몰렸다. LH...

공급 물량에 비해 엄청난 경쟁률

- 당해 공고의 매물량이 많지 않고
- 주택공사에서 따로 제공하는 청약 지표가 없어서
- 어느 지역, 어떤 단지에 청약을 넣어야 할지 판단이 어려운 상황

2022년 2차 청년 매입임대주택 입주자모집공고[2022.12.30.]
인터넷 청약신청 최종 경쟁률 공지

이투데이 | 2023.04.25.

“보증된 집에 살자”...올해 수도권 LH 청년매입임대 경쟁률 '45대 ...

이투데이=정용욱 기자 | 전세사기 사태 확산 이후 공공 임대주택 선호 급증 올해 수도권 매입임대 주택 인기가 천정부지로 치솟고 있다. 청년 매입임대주택 경쟁...

S 시사저널이코노미 | 2023.04.16.

432가구 모집에 4만명 신청...공공임대에 몰리는 청년들

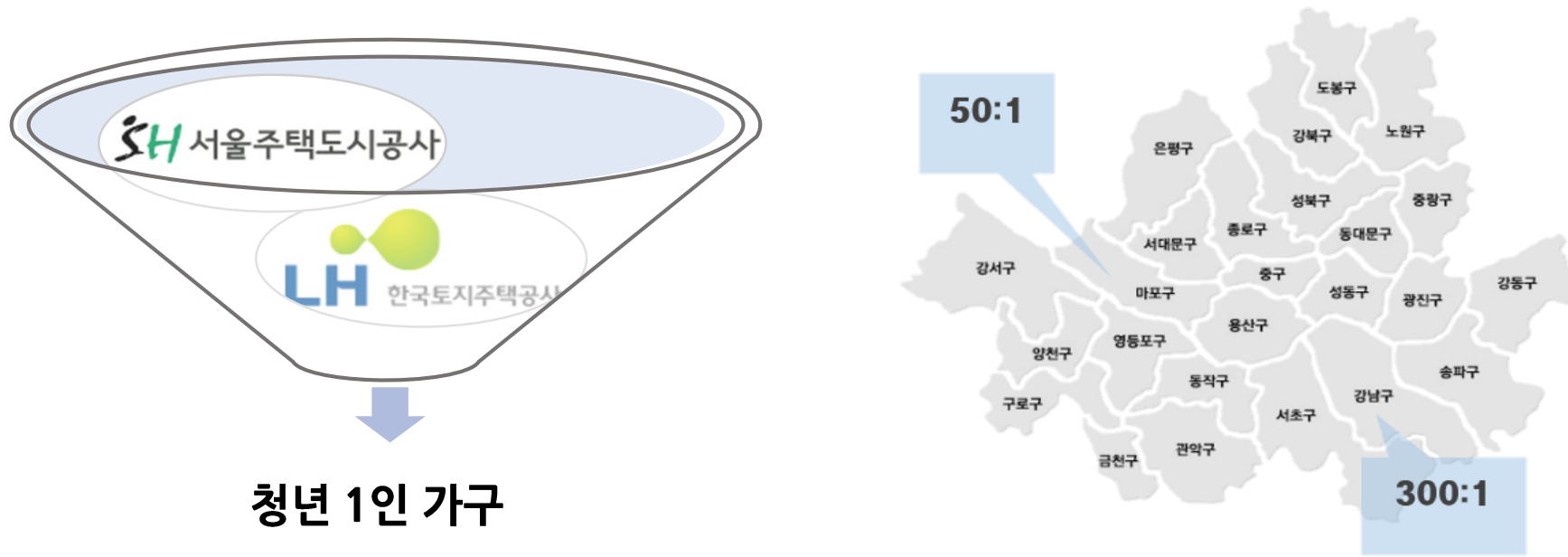
/ 사진=연합뉴스 수도권 공공임대주택이 청년들로부터 인기를 끌고 있다. 빌라왕 전세사기 여파로 보증금이 높은 주택을 꺼리는 현상이 심화된 데다 '전세의 월세...

공급호수(호)	신청현황(누계)	
	신청자(명)	경쟁률
2	62	31
2	22	11
2	12	6
2	9	4.5
5	159	31.8
1	115	115
1	92	92
1	174	174
1	120	120
1	131	131
1	113	113
2	200	100
8	631	78.9
2	108	54
2	420	210

*출처 :sh주택공사

서울시 1인 가구 임대주택 경쟁률 예측 서비스

- ✓ 비싼 월세에 부담을 느끼는 1인가구 청년
- ✓ 치열한 임대주택 청약 경쟁률 속에서 청약 신청의 지표가 필요한 청년들
- 서울시 지역별 LH, SH 청년 임대주택 경쟁률 분석 후 예측하여 시각화



지역별 청약 **커트라인** 예측 서비스

- 사용자의 순위, 가산점과 지역별 예측 커트라인을 비교하여 제공
- 당해 공고 선호 지역별 사용자 수를 실시간 제공
- 청약 성공 확률을 높일 수 있음



현재 50명이 마포구를 선호하고 있습니다.
마포구 선호 비율은 30%입니다.



사용자님의 가산점보다
커트라인이 낮은 지역은
마포구, 구로구, 금천구입니다.



마포구 예측 경쟁률은 50:1입니다.
마포구는 지역별 예측 경쟁률 2위입니다.

Language



1.8.0_371



개발 환경



4.3.9



2019.09



11.2.0.2



4.3.1

Server

Apache
Tomcat
9.0

2

기술 stacks



- java1.8 기반 미니 프로젝트
- 은행 입출금 시스템
- 주소록 관리 시스템
- 도서 대여 프로그램



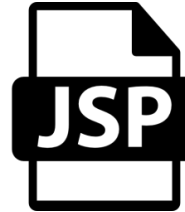
- Spring 4.3.9 version
- 전자정부 프레임워크3.9 에서 Spring Framework를 이용해 CRUD 를 구현할 수 있음
- 게시판 MVC2패턴



- 프로젝트 UI 제작
- bootstrap 반응형 페이지 제작



- 회원관리 유효성 검사
- 비동기 통신(AJAX)을 구현
- chart.js



- MVC2방식의 임대주택 청약 예측 서비스 웹사이트 구현
- 회원관리 시스템(등록, 삭제, 수정)
- 게시판 (등록, 삭제, 페이징 처리)



- 데이터 베이스 설계 (ERD)
- JOIN, Group by 등과 같은 쿼리문 작성
- 게시판 / 회원관리 등의 테이블 관리



- 데이터 분석
- 회귀 모델로 예측



- 기본적인 문법 학습



- Ubuntu 기반 Linux개발 환경 구축 실습

개발 스케줄표

[illegible]

서울시 1인 가구 임대주택 지역별 청약 **커트라인**, **경쟁률** 예측 데이터 분석

수집 → 전처리 → 분석 + 예측(모델링) → db

 서울주택도시공사

 행정안전부

 Pay 부동산



지역별 경쟁률, 커트라인 예측 데이터 수집

✓ SH 주택 공사 제공하는 기본 데이터

- **rate** : 경쟁률
- **size** : 평형
- **supply** : 공급 세대수
- **value** : 임대주택 월세 (보증금 3천만원 기준)
- **people** : 매물별 지원자 수
- **cut** : 청약 커트라인

서울시 1인가구 실태조사에 따르면 청년 1인가구 밀집지역은 학교, 직장근처 변화가 접근성, 편의시설에 영향을 많이 받음.

✓ 네이버 부동산 제공

- **d.sub** : 지하철 역까지의 거리(네이버 부동산 제공),
- **d.hos** : 병원까지의 거리

지역별 특성을 반영하기 위해

✓ 행안부 제공

- **security** : 지역별 사회안전지수(경제활동, 생활안전, 건강보건, 주거환경의 지표 반영),
- **avg.price** : 지역별 평균 집값

✓ 경쟁률 예측 독립변수

- x1 : size
- x2 : supply
- x3 : value
- x4 : d.sub
- x5 : d.hos
- x6 : security
- x7 : avg.price
- x8 : people

✓ 경쟁률 예측 종속변수

- y : rate

✓ 커트라인 예측 독립변수

- x1 ~ x7 동일
- x8 : rate

✓ 커트라인 예측 종속변수

- y : cut

지역별 경쟁률 예측 모델링 및 결과 분석

```
> model1 <- lm(rate ~ size + supply + value + d.sub + d.hos + security
+               + avg.price + people, data = testset)
> summary(model1)
```

```
Call:
lm(formula = rate ~ size + supply + value + d.sub + d.hos + security +
    avg.price + people, data = testset)
```

Residuals:

Min	1Q	Median	3Q	Max
-346.82	-38.55	-15.46	12.11	714.70

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.829e+02	8.601e+01	3.289	0.00103 **
size	2.896e+00	4.045e-01	7.160	1.41e-12 ***
supply	-9.933e-01	1.140e-01	-8.713	< 2e-16 ***
value	-3.177e-04	4.035e-05	-7.873	7.77e-15 ***
d.sub	-3.897e-02	9.591e-03	-4.064	5.15e-05 ***
d.hos	-8.663e-03	1.138e-02	-0.761	0.44681
security	-5.817e+00	1.701e+00	-3.419	0.00065 ***
avg.price	2.063e-02	3.190e-03	6.467	1.46e-10 ***
people	1.892e-01	7.282e-03	25.988	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 91.67 on 1185 degrees of freedom
Multiple R-squared: 0.4211, Adjusted R-squared: 0.4172
F-statistic: 107.8 on 8 and 1185 DF, p-value: < 2.2e-16

✓ 다중선형 회귀 모델

- 여러개의 독립변수를 다루는 다중선형 회귀 모델 사용

✓ 결측치 처리

- 평균값 대체

✓ 회귀식

- $\text{rate} = 282.9 + 2.896 \times \text{size} - 0.9933 \times \text{supply} - 0.0003177 \times \text{value} - 0.03897 \times \text{d.sub} - 0.008663 \times \text{d.hos} - 5.817 \times \text{security} + 0.02063 \times \text{avg.price} + 0.1892 \times \text{people}$

✓ p-value

- 2.2×10^{-16} 으로 신뢰수준이 95% 이상임을 나타냄.

✓ Adjust R-squared

- 모델의 설명력을 나타내며 0.4172 로 다소 낮은값.
현실을 41% 정도 설명할수 있음.

지역별 커트라인 예측 데이터 전처리 및 모델링

```
# 결측치 처리 (중앙값으로 대체)
data$size <- ifelse(is.na(data$size), median(data$size, na.rm = TRUE), data$size)
data$people <- ifelse(is.na(data$people), median(data$people, na.rm = TRUE), data$people)
data$rate <- ifelse(is.na(data$rate), median(data$rate, na.rm = TRUE), data$rate)
data$value <- ifelse(is.na(data$value), median(data$value, na.rm = TRUE), data$value)
data$d.sub <- ifelse(is.na(data$d.sub), median(data$d.sub, na.rm = TRUE), data$d.sub)
data$security <- ifelse(is.na(data$security), median(data$security, na.rm = TRUE),
data$security)
data$avg.price <- ifelse(is.na(data$avg.price), median(data$avg.price, na.rm = TRUE),
data$avg.price)
data$supply <- ifelse(is.na(data$supply), median(data$supply, na.rm = TRUE), data$supply)
data$cut <- ifelse(is.na(data$cut), median(data$cut, na.rm = TRUE), data$cut)
data <- na.omit(data)

for (col in names(trainData)) {
  trainData[[col]][is.na(trainData[[col]])] <- mean(trainData[[col]], na.rm = TRUE)
}

for (col in names(testData)) {
  testData[[col]][is.na(testData[[col]])] <- mean(testData[[col]], na.rm = TRUE)
}

# 데이터 분할
set.seed(42)
trainIndex <- sample(1:nrow(data), 0.8*nrow(data))
trainData <- data[trainIndex,]
testData <- data[-trainIndex,]

# 랜덤 포레스트 모델 학습
rf_model <- randomForest(cut ~ d.hos + size + rate + value + d.sub + security + avg
.price + supply, data=trainData, ntree=100)
```

✓ 랜덤 포레스트

- 지역별 청약 커트라인은 연속성 데이터가 아니기 때문에

랜덤포레스트를 선택

- 수백 개의 의사결정나무를 결합하여 이를 평균화한 모형을 생성하는

앙상블 학습 방법

✓ 전처리

- 데이터는 80%의 훈련 데이터와 20%의 테스트 데이터로 분할

- 결측치 처리 (중앙값으로 대체) 및 제거

- 훈련 데이터의 결측치 처리 (평균값으로 대체)

- 테스트 데이터의 결측치 처리 (평균값으로 대체)

✓ 100개의 트리를 사용하여 모델을 학습

지역별 **커트라인** 예측 결과 분석 및 한계

```
# 예측
predictions <- predict(rf_model, newdata=testData)

results <- data.frame(region = testData$region, Predicted = predictions)

# 지역별 예측값의 평균
region_avg_predictions <- aggregate(Predicted ~ region, data = results, FUN = mean)

> # 결과 출력
> print(region_avg_predictions)
  region Predicted
1  강남구  25.30619
2  강동구  20.83698
3  강북구  19.97523
4  강서구  24.31421
5  관악구  22.24189
6  광진구  25.55644
7  구로구  15.38101
8  금천구  17.34525
9  노원구  24.82833
10 도봉구  20.80770
11 동대문구 24.71734
12 동작구  22.70767
13 마포구  25.76011
14 서대문구 22.30989
15 서초구  25.74415
16 성동구  22.50904
17 성북구  22.45666
18 송파구  22.90332
19 양천구  17.24008
20 영등포구 22.28743
21 용산구  26.41230
22 은평구  17.70094
23 종로구  27.05767
24 중구    25.39868
25 중랑구  20.65259

> # 성능 평가 (RMSE)
> rmse <- sqrt(mean((predictions - testData$cut)^2))
> print(rmse)
[1] 1.679401

> mae <- mean(abs(predictions - testData$cut))
> # 결과 출력
> print(paste("MAE:", mae))
[1] "MAE: 1.13411872009695"

> # 특성 중요도 확인
> importance(rf_model)
              IncNodePurity
d.hos           3735.009
size            2865.574
rate            15395.639
value           3277.473
d.sub           4676.697
security        1987.411
avg.price       2428.313
supply          2181.874
```

✓ 성능 평가

- RMSE는 예측 오차의 제곱합의 평균의 제곱근으로, 값이 작을수록 예측의 정확도가 높다는 것을 의미.
- MAE (Mean Absolute Error): MAE는 실제 값과 예측 값의 차이를 절댓값으로 변환하여 평균한 것으로, 값이 작을수록 예측의 정확도가 높다는 것을 의미.

✓ 예측 변수의 중요도 지수

- 모델의 예측 변수별 중요도를 확인한 결과, **경쟁률**이 가장 높은 중요도를 보였으며 그 뒤로 지하철역까지의 거리, 병원까지의 거리 순서로 중요도가 높게 나타남.

✓ 한계

- 데이터셋의 수가 1185개로 부족함
 - 데이터가 있는 예측변수의 수가 부족함
- 청약가점 및 경쟁률에 영향을 미치는 다양한 외부 요인들이 더 있을 수 있으며, 완전한 결과를 도출하는 데 한계가 있음.

3. 기능적 요구사항

3.1 청약 진단하기 기능

3.1.1 아이디 입력

- 사용자는 자신의 아이디를 텍스트 형식으로 입력할 수 있어야 한다.
- "아이디를 입력하세요" 라는 안내 메시지와 함께 입력 폼 제공

3.1.2 선호 지역 선택

- 사용자는 드롭다운 메뉴에서 선호 지역을 선택할 수 있어야 한다.
- 마포구, 강남구, 강서구 등의 서울시 자치구 옵션 제공

3.1.3 순위 선택

- 사용자는 자신의 순위(1 순위, 2 순위, 3 순위) 중 하나를 버튼을 통해 선택할 수 있어야 한다.
- 각 순위에 대한 상세 설명 제공

3.1.4 가산점 선택

- 사용자는 자신에게 해당되는 가산점 항목을 체크박스를 통해 선택할 수 있어야 한다.
- 각 가점에 대한 설명 제공하며, 복수 선택 가능

3.1.5 결과 제출

- 사용자는 입력된 정보를 바탕으로 가점을 계산하고 결과값을 얻기 위해 폼을 제출할 수 있어야 한다.

3.2 청약 예측 및 분석 기능

- 사용자가 입력한 순위와 가점을 바탕으로 청약 예측정보를 제공할 수 있어야 한다.

3.2.1 선호지역 카드

- 사용자의 선호지역을 표시한다.

3.2.2 선호지역 예측 경쟁률 카드

- 사용자가 입력한 선호지역에 대한 예측 경쟁률을 보여줄 수 있어야 한다.

3.2.3 선호지역 예측 커트라인 카드

- 사용자가 입력한 선호지역에 대한 예측 커트라인을 보여줄 수 있어야 한다.
- 커트라인 점수에 따른 프로그레스 바를 통해 시각적으로 표시한다.

3.2.4 월별 경쟁률 차트

- 사용자의 선호지역의 월별 실제 경쟁률을 차트로 시각화하여 표시한다.

3.2.5 지역별 경쟁률 순위 파이 차트

- 각 지역의 경쟁률 상위 5 개 지역을 파이 차트로 시각화하여 표시한다.

3.2.6 사용자의 가정보다 예측 커트라인이 낮은 지역 리스트

- 사용자의 순위와 가산점보다 예측 커트라인이 낮은 지역과 해당지역의 커트라인을 리스트 형태로 모두 출력하여 표시한다.
- 각 지역들의 예측 커트라인을 프로그레스 바를 통해 시각적으로 표시한다.

3.2.7 페이지 이동

- "처음으로" 버튼을 통해 가산점 계산하기 페이지로 이동할 수 있어야 한다.

3.3 메인페이지

3.3.1 네비게이션 바

- 홈, 모집공고, 청약진단하기, 상세조회, ABOUT 등의 메뉴를 포함한다.
- 청약 진단하기 및 상세조회 메뉴는 드롭다운 형태로 하위 메뉴를 포함한다.
- 검색 모달이 있으며 버튼 클릭 시 사용자가 검색어를 입력할 수 있다.

3.3.2 캐러셀 기능

- 캐러셀 영역에 이미지 전환과 함께 멘트를 표시할 수 있어야 한다.

3.3.3 지도 조회 기능

- 지도상의 자치구 영역에 마우스 호버 시 예측 경쟁률 박스로 표시할 수 있어야 한다.

3.3.4 매물 조회 기능

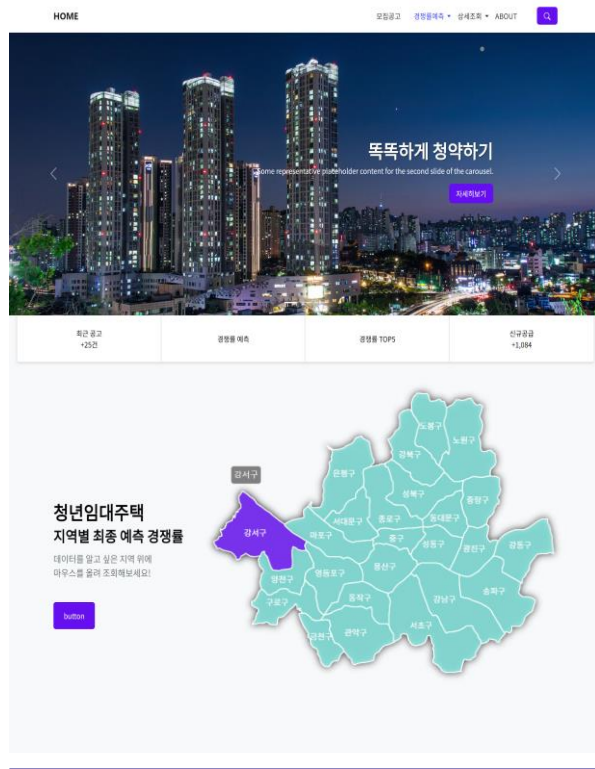
- 사용자가 지역을 선택하면 해당지역의 경쟁률이 높았던 단지 5 개를 카드형태의 캐러셀로 표시해 보여줄 수 있어야 한다.

3.3.5 서비스 이동 기능

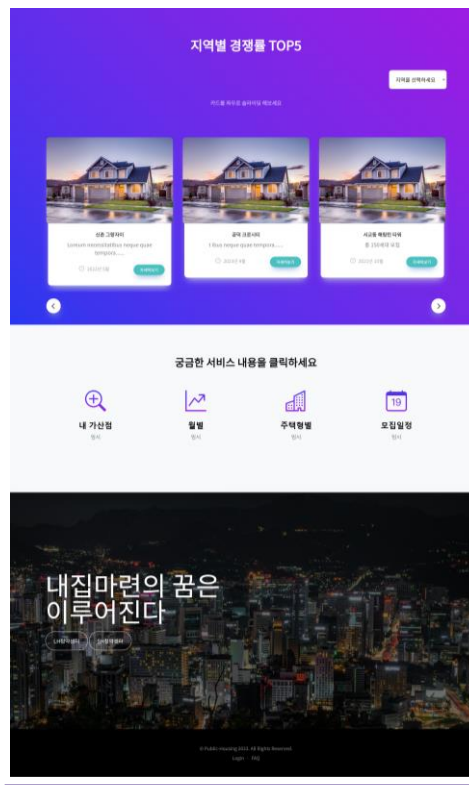
- 각 기능을 클릭하면, 해당 기능과 관련된 페이지나 섹션으로 이동 할 수 있어야 한다.

6

화면 개요



메인 페이지



메인 페이지



청약 진단



예측 리포트

6

화면 설계서

화면 이름	메인 페이지	화면 경로	메인 페이지	화면 ID	-	작성일	2023/05/30
<div><div><div><div><div><div>←</div><div>→</div><div>↺</div></div><div>HOME</div><div><div>1</div><div>모집공고</div></div><div><div>2</div><div>청약 진단하기</div><div>상세조회</div><div>ABOUT</div></div><div><div>3</div><div>4</div></div><div><div>5</div><div>최근 공고 +25건</div><div>경쟁률 예측</div><div>카드라인 예측</div><div>신규공급 +500</div></div></div><div><div><div><div><div>1</div><div>임대주택 정보 메뉴 네비게이션</div></div><div><div>2</div><div>메뉴 <u>드롭다운</u> 형태로 <u>하위메뉴</u> 포함</div></div><div><div>3</div><div>메인 화면 캐로셀</div></div><div><div>4</div><div>캐로셀 인디케이터(화살표)</div></div><div><div>5</div><div>업데이트 내역 표시</div></div></div></div></div></div></div></div>							

화면 이름	상세 조회 페이지	화면 경로	메인 페이지> 지역별 경쟁률 TOP5	화면 ID	-	작성일	2023/05/30
Description							
1	제목 표시						
2	지역 선택 드롭 다운						
3	슬라이딩 동작 안내 메시지						
4	해당 지역 경쟁률 상위 5개 단지 카드 캐로셀 형태로 표시						
5	카드 캐로셀 인디케이터						

화면 이름	경쟁률 예측 페이지	화면 경로	메인 페이지 > 경쟁률 예측 페이지	화면 ID	-	작성일	2023/05/30
-------	------------	-------	---------------------	-------	---	-----	------------

HOME

모집공고청약 진단하기상세조회ABOUT

🔍4

1

청년임대주택
지역별 최종 예측 경쟁률

데이터를 알고 싶은 지역 위에
마우스를 올려 조회해보세요!

button

2

3

15:1

강서구

강서구강

화면 이름	메인 페이지	화면 경로	메인 페이지 > 서비스 내용 색인	화면 ID	-	작성일	2023/05/30
Description							
1	제목 표시						
2	아이콘으로 시각화						
3	각종 서비스 내용 표시						
4	서비스 내용에 대한 설명						

화면 설계서

화면 이름	청약 진단 페이지	화면 경로	청약 진단하기 > 내 가점 계산하기	화면 ID	-	작성일	2023/05/31
-------	-----------	-------	---------------------	-------	---	-----	------------

The screenshot displays the '내 가점 계산하기' (My Points Calculation) interface. It features a top navigation bar with tabs for 'HOME', '모집공고', '청약 진단하기', '상세조회', and 'ABOUT'. Below the navigation bar, there are four numbered steps indicating the process flow:

- Step 1:** Introduction to point calculation rules. It includes two main criteria:
 - Family status and household registration points (e.g., 100% for single/family, 50% for married).
 - Income-based points (e.g., 100% for income below 36,100,000 KRW, 50% for higher income).
- Step 2:** Partially visible, showing rules related to housing type.
- Step 3:** Titled '다음 가산점 항목을 해당하는 항목을 모두 클릭하세요(복수선택):' (Click all applicable items for the next bonus points). It lists several categories for selection:
 - Birth Order (e.g., 1st child gets 1 point)
 - Educational Level (e.g., University graduate gets 1 point)
 - Housing Type (e.g., Rental gets 1 point)
 - Other factors like age, disability, etc.
- Step 4:** Labeled '완료하기' (Complete), highlighted with a red circle.

[illegible]

